SECTION III

BOVINE CORNEAL OPACITY AND PERMEABILITY (BCOP) TEST METHOD ACCURACY AND RELIABILITY REANALYSIS

[This Page Intentionally Left Blank]

1.0 INTRODUCTION

On November 1, 2004, NICEATM released draft BRDs on the current status of four *in vitro* test methods for detecting ocular corrosives and severe irritants (see http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm). The test methods reviewed were the BCOP, the HET-CAM, the IRE, and the ICE assays. On January 11-12, 2005, ICCVAM convened an Expert Panel to independently evaluate the validation status of the four *in vitro* test methods for identifying ocular corrosives or severe irritants. The Expert Panel Report, *Evaluation of the Current Validation Status of In Vitro Test Methods for Identifying Ocular Corrosives and Severe Irritants*, can be obtained by contacting NICEATM or electronically from http://iccvam.niehs.nih.gov/methods/eyeirrit.htm. Public comments at the meeting revealed that additional data could be made available that had not yet been provided in response to earlier requests for data. The Expert Panel subsequently recommended that the additional data be requested and that a reanalysis of the accuracy and reliability of each test method be conducted, to the extent possible.

In response to this recommendation, a second *FR* notice was published on February 28, 2005 (*FR* Vol. 70, No. 38, pp. 9661-9662; http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) requesting all available *in vitro* data on these four *in vitro* ocular irritancy test methods and corresponding *in vivo* rabbit eye test method data, as well as any human exposure data (either via ethical human studies or accidental exposure). The first *FR* notice requesting these data had been published on March 24, 2004 (*FR* Vol. 69, No. 57, pp. 13859-13861; http://iccvam.niehs.nih.gov/methods/eyeirrit.htm). Also, a request for relevant data was re-sent directly to the primary developers or users of each test method, and sent to other scientists who participated in or attended the Expert Panel Meeting on January 11-12, 2005 and who had indicated a desire to provide additional data. No human exposure data was obtained for the substances evaluated in the BCOP test method, and therefore no calculations could be made on the accuracy of the BCOP test method for predicting human severe ocular irritancy.

Other factors also necessitated a reanalysis of the accuracy of BCOP for detecting ocular corrosives and severe irritants. First, clarification regarding the rules for classification of severe irritants was obtained subsequent to the release of the four BRDs that resulted in changes to the hazard classification of some of the substances used in the original analysis. For the original analysis, reversibility of ocular effects for all EU (EU [2001]) and GHS ([UN 2003]) hazard classification systems was considered to be achieved if, by post-exposure day 21, the endpoint scores fell below the threshold that resulted in a test substance being classified as a severe irritant. The new information obtained indicated that reversibility of ocular effects is achieved only when all scores reach zero by post-exposure day 21. This change resulted in a small number of substances previously classified as non-severe irritants now being classified as severe irritants.

Second, the chemical classes assigned to each test substance were revised to reflect a standardized classification scheme (based on MeSH; [http://www.nlm.nih.gov/mesh]) that would ensure consistency in classifying substances among all *in vitro* ocular test methods under consideration. This resulted in some chemicals being re-classified into other chemical

classes. The accuracy of the BCOP test method, by chemical class and using the GHS classification system (UN [2003]), has been reanalyzed to reflect these changes.

Finally, an additional accuracy analysis was conducted. In this analysis, the accuracy of each *in vitro* ocular irritancy test method for detecting ocular corrosives or severe irritants, depending on whether the classification was based on the severity of the response and/or its persistence to day 21 post-treatment, was determined.

For the BCOP test method, the changes to the existing database that resulted from using the appropriate persistence classification criteria and any new data and/or information received subsequent to the release of the draft BRD are summarized in Table III-1. At the Expert Panel meeting, the *in vivo* rabbit eye test data that corresponded to the substances tested in BCOP in the Gautheron et al. (1994) study were received from Johnson & Johnson Pharmaceutical R&D. Individual cornea data from the BCOP tests evaluating these 52 substances also were provided subsequent to the meeting. Johnson & Johnson Pharmaceutical R&D also provided individual cornea data for 20 substances evaluated in the BCOP test method, comparing results achieved using corneas from adult animals (>24 months) versus those from young animals (6 - 8 months). The efforts of Drs. Freddy Van Goethem and Philippe Vanparys that provided this additional data are gratefully acknowledged.

2.0 ACCURACY OF THE BCOP TEST METHOD – REANLYSIS

The ability of the BCOP test method to correctly identify ocular corrosives and severe irritants, as defined by the GHS, EPA, and EU classification systems (EPA [1996]; EU [2001]; UN [2003])[1], was evaluated. The three regulatory ocular hazard classification systems considered during this analysis use different classification systems and decision criteria to identify ocular corrosives and severe irritants based on *in vivo* rabbit eye test results. All three classification systems are based on individual animal data in terms of the magnitude of the response and on the extent to which induced ocular lesions fail to reverse by day 21. However, there are differences among the three classifications systems with regard to the criteria used by NICEATM for distinguishing between a severe and a nonsevere response (See **Appendix A**). Thus, to evaluate the accuracy of the HET-CAM test method for identifying ocular corrosives and severe irritants, individual rabbit data collected at the different observation times was needed for each substance.

The ability of the BCOP test method to correctly identify ocular corrosives and severe irritants, as defined by the GHS, EPA, and EU classification systems (EPA [1996]; EU [2001]; UN [2003]), was evaluated using two approaches. In the first approach, the accuracy of BCOP was assessed separately for each *in vitro-in vivo* comparative study (i.e., publication) reviewed in Sections 4.0 and 5.0 of the draft BCOP BRD. In the second approach, an overall analysis of BCOP test method accuracy was conducted by combining

[1] For the purposes of this analysis, an ocular corrosive or severe irritant was defined as a substance that would be classified as Category 1 according to the GHS classification system (UN [2003]), as Category I according to the EPA classification system (EPA [1996]), or as R41 according to the EU classification system (EU [2001]).

1882 **Table III-1. Summary of BCOP Database Changes**
1883

Data Source	Data Base	Number of Available Substances	Number of Acceptable Substances by Ocular Irritancy Classification System			*Comments*
			EPA[1]	EU[2]	GHS[3]	
			Cat[4] I/Total[5]	R41/Total	Cat 1/Total	
Gautheron (1994)	New[6]	51	7/48	7/48	7/47	Additional *in vivo* animal data were received subsequent to the original analysis that allowed for classification according to all three classification systems.
	Old[6]	51	6/12	8/51	7/13	
Balls et al. (1995)	New	59	18/53	19/50	22/54	The decrease in the total number of usable substances is due to excluding substances from consideration due to insufficient *in vivo* rabbit eye test data for classification (See **Appendix A**).
	Old	59	20/55	21/59	22/57	
Swanson et al. (1995)	New	20	6/8	6/9	6/8	The decrease in the total number of usable substances is due to excluding substances from consideration due to insufficient *in vivo* rabbit eye test data for classification (See **Appendix A**). The increase in the number of corrosive/severe irritants is due to the reclassification of substances.
	Old	20	6/9	5/9	6/9	
Casterton (1996)	New	97	27/56	25/54	27/55	The decrease in the total number of usable substances is due to excluding substances from consideration due to insufficient *in vivo* rabbit eye test data for classification (See **Appendix A**). The increase in the number of corrosive/severe irritants is due to the reclassification of substances.
	Old	97	26/55	24/60	26/56	
Gettings (1996)	New	25	10/25	8/23	8/23	The decrease in the total number of usable substances is due to excluding substances from consideration due to insufficient *in vivo* rabbit eye test data for classification
	Old	25	10/25	6/25	8/25	

Data Source	Data Base	Number of Available Substances	Number of Acceptable Substances by Ocular Irritancy Classification System			*Comments*
			EPA[1]	EU[2]	GHS[3]	
			Cat[4] I/Total[5]	R41/Total	Cat 1/Total	
						(See **Appendix A**). The increase in the number of corrosive/severe irritants is due to the reclassification of substances.
Southee (1998)	New	16	5/14	6/14	7/15	The decrease in the total number of usable substances is due to excluding substances from consideration due to insufficient *in vivo* rabbit eye test data for classification (See **Appendix A**). The change in the number of corrosive/severe irritants is due to the reclassification of substances.
	Old	16	6/14	5/15	6/14	
Swanson and Harbell (2000)	New	13	4/9	1/9	1/9	
	Old	13	4/9	1/9	1/9	
Bailey (2004)	New	16	1/13	3/13	3/14	The decrease in the total number of usable substances is due to excluding substances from consideration due to insufficient *in vivo* rabbit eye test data for classification (See **Appendix A**). The change in the number of corrosive/severe irritants is due to the reclassification of substances.
	Old	16	3/16	3/16	3/16	

[1]EPA = U.S. Environmental Protection Agency (EPA [1996]).
[2]EU = European Union (EU [2001]).
[3]GHS = Globally Harmonized System (UN [2003]).
[4]Cat = Category.
[5]First number (before forward slash) refers to the number of substances in each study that were classified as a severe irritant according to each classification system (EPA, EU, and GHS). The second number (after the forward slash) refers to the number of substances in were classified, based on animal data, for each classification system (EPA, EU, GHS).
[6]New = accuracy statistics based on revised analysis; New = accuracy statistics based on the previous analysis included in the draft BCOP BRD.

results from each study, and then assigning an overall ocular irritancy classification for each substance. When the same substance was evaluated in multiple laboratories, the overall BCOP ocular irritancy classification was based on the majority of calls among all of the studies. When there was an even number of different irritancy classifications for substances (e.g., two tests classified a substance as a nonsevere irritant and two tests classified a substance as a severe irritant), the more severe irritancy classification was used for the overall classification for the substance (severe irritant, in this case).

Based on the revisions made to the BCOP test method database, a revised accuracy analysis has been conducted. The calculations were performed as described previously in Section 6.0 of the draft BRD. To allow for a comparison of the results obtained in the revised analysis relative to those obtained previously, the data tables below include accuracy statistics from both analyses. However, the discussion of the results in the sections that follow relate to the revised analysis only.

2.1 GHS Ocular Hazard Classification System

The eight studies (Gautheron et al. [1994]; Balls et al. [1995]; Swanson et al. [1995]; Casterton et al. [1996]; Gettings et al. [1996]; Southee [1998]; Swanson and Harbell [2000]; Bailey et al. [2004]) contained BCOP test data on 203 substances, 161 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification as defined by the GHS classification system (UN [2003])[2] (see **Appendix III-A**). Based on results from *in vivo* rabbit eye experiments, 53[3] of the 161 substances were classified as severe irritants (i.e., Category 1), the other 108 substances were classified as nonsevere irritants (either Category 2A, 2B) or nonirritants. The 42 substances that could not be classified according to the GHS classification system due to the lack of adequate animal data are so noted in **Appendix III-A**.

Based on the data provided in the eight studies, and when results across multiply tested substances were combined to generate a single consensus call per test substance, the BCOP test method has an accuracy of 70% to 93%, a sensitivity of 57% to 100%, a specificity of 66% to 100%, a false positive rate of 0% to 34%, and a false negative rate of 0% to 52% (**Table III-2**).

In terms of an overall accuracy analysis, combining the data from Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000), and Bailey et al. (2004), the BCOP test method has an accuracy of 81% (119/147), a sensitivity of 84% (36/43), a specificity of 80% (83/104), a false positive rate of 20% (21/104), and a false negative rate of 16% (7/43). The performance characteristics for the pooled studies are provided in **Table III-2**. Similar to the original accuracy analysis, data from Casterton et al. (1996) were not included in the overall accuracy analysis since the

[2] For the purpose of this accuracy analysis, *in vivo* rabbit study results were used to identify GHS Category 1 irritants (i.e., severe irritants); substances classified as GHS Category 2A and 2B irritants were identified as nonsevere irritants.

[3] One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice in the same laboratory. The results were discordant with respect to GHS classification. According to one test, the classification was Category 1, while results from the other test yielded a Category 2A classification. The accuracy analysis was performed with the substance classified as Category 1.

Table III-2. Evaluation of the Performance of the BCOP Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the GHS[1] Classification System, by Study and Overall

Data Source	Anal.[2]	N[3]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[4]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Gautheron et al. 1994 (new)[5]	IVIS	47/52	74[6]	35/47	71	5/7	75	30/40	33	5/15	94	30/32	25	11/40	29	2/7
Gautheron et al. 1994 (old)[5]	IVIS	13/52	77[6]	10/13	71	5/7	83	5/6	83	5/6	71	5/7	17	1/6	29	2/7
Balls et al. 1995 (new)[7]	IVIS	54/59	70[6]	38/54	77	17/22	66	21/32	61	17/28	81	21/26	34	11/32	23	5/22
Balls et al. 1995 (old)	IVIS	57/59	70[6]	40/57	77	17/22	66	23/35	59	17/29	82	23/28	34	12/35	23	5/22
Swanson et al. 1995 (new)	IVIS	8/20	100	8/8	100	6/6	100	2/2	100	6/6	100	2/2	0	0/2	0	0/6
Swanson et al. 1995 (old)	IVIS	9/20	89	8/9	100	6/6	67	2/3	86	6/7	100	2/2	33	1/3	0	0/6
Gettings et al. 1996 (new)	Perm	23/25	87	20/23	75	6/8	93	14/15	86	6/7	88	14/16	7	1/15	25	2/8
Gettings et al. 1996 (old)	Perm	25/25	88	22/25	75	6/8	94	16/17	86	6/7	89	16/18	6	1/17	25	2/8
Casterton et al. 1996 (new)	O/P	55/97	67	37/55	48	13/27	86	24/48	76	13/17	63	24/38	14	4/28	52	14/27
Casterton et al. 1996 (old)	O/P	56/97	66	37/56	46	12/26	83	25/30	71	12/17	64	25/39	17	5/30	54	14/26
Southee 1998 (new)	IVIS	15/16	73	11/15	57	4/7	88	7/8	80	4/5	70	7/10	12	1/8	43	3/7
Southee 1998 (old)	IVIS	14/16	64[6]	9/14	50	3/6	75	6/8	40	2/5	67	6/9	25	2/8	50	3/6
Swanson & Harbell 2000 (new)	IVIS	9/13	78	7/9	100	1/1	75	6/8	33	1/3	100	6/6	25	2/8	0	0/1
Swanson & Harbell 2000 (old)	IVIS	9/13	78	7/9	100	1/1	75	6/8	33	1/3	100	6/6	25	2/8	0	0/1

Data Source	Anal[2]	N[3]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[4]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Bailey et al. 2004 (new)	IVIS	14/16	93	13/14	67	2/3	100	11/11	100	2/2	92	11/12	0	0/11	33	1/3
Bailey et al. 2004 (old)	IVIS	16/16	94	15/16	67	2/3	100	13/13	100	2/2	93	13/14	0	0/13	33	1/3
Entire Data Set[8] (new)		147/203	81	119/147	84	36/43	80	83/104	63	36/57	92	83/90	20	21/104	16	7/43
Entire Data Set (old)		120/200	79	95/120	76	32/42	81	63/78	69	34/49	86	61/71	19	15/78	24	10/42

[1]GHS = Globally Harmonized System (UN [2003]).

[2]Anal. = Analytical method used to transform the sample data into BCOP classification. IVIS = *In Vitro* Irritancy Score developed by Gautheron et al. (1994). Perm = Permeability value only used to classify *in vitro* ocular irritancy in the BCOP assay; an OD_{490} value >0.600 was considered a severe irritant. O/P = irritation class based on the endpoint (opacity or permeability) with the highest score for its respective range (Casterton et al. [1996]).

[3]N = Number of substances included in this analysis/the total number of substances evaluated in the study.

[4]Data used to calculate the percentage.

[5]New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the previous analysis included in the draft BCOP BRD.

[6]Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the multiple testing laboratories and tests (for substances tested multiple times in a laboratory).

[7]The test substance 1% benzalkonium chloride was tested in two different *in vivo* studies, producing discordant results with respect to GHS classification; the analysis was performed using the Category 1 classification.

[8]Data from Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000), and Bailey et al. (2004) were pooled together and an overall *in vitro* classification was assigned for each test substance based on the majority and/or most severe classification obtained across tests and testing laboratories. Data from Casterton et al. (1996) were not included in this analysis since the protocol used to generate BCOP data differed considerably from the other studies (e.g., a spectrophotometer was used to measure opacity instead of an opacitometer, and solids were applied neat instead of as a 20% solution or suspension).

protocol used to generate BCOP data differed considerably from the other studies (e.g., a spectrophotometer was used to measure opacity instead of an opacitometer; solids were applied neat instead of as a 20% solution or suspension).

2.2 EPA Ocular Hazard Classification System

The eight studies (Gautheron et al. [1994]; Balls et al. [1995]; Swanson et al. [1995]; Casterton et al. [1996]; Gettings et al. [1996]; Southee [1998]; Swanson and Harbell [2000]; Bailey et al. [2004]) contained BCOP test data on 203 substances, 160 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification as defined by the EPA classification system (EPA [1996])[4] (see **Appendix III-A**). Based on results from *in vivo* rabbit eye experiments, 50 of the 160 substances were classified as severe irritants (i.e., Category I), the other 110 substances were classified as nonsevere irritants (either Category II, III, or IV). The 43 substances that could not be classified according to the EPA classification system due to the lack of adequate animal data are so noted in **Appendix III-A**.

Based on the data provided in the eight studies, and when results across multiply tested substances were combined to generate a single consensus call per test substance, the BCOP test method has an accuracy of 62% to 92%, a sensitivity of 0% to 100%, a specificity of 50% to 100%, a false positive rate of 0% to 50%, and a false negative rate of 0% to 100% (**Table III-3**).

In terms of an overall accuracy analysis, combining the data from Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000), and Bailey et al. (2004), the BCOP test method has an accuracy of 79% (113/143), a sensitivity of 75% (30/40), a specificity of 81% (83/103), a false positive rate of 19% (20/103), and a false negative rate of 25% (10/40). The performance characteristics for the pooled studies are provided in **Table III-3**. Data from Casterton et al. (1996) were not included in the overall accuracy analysis since the protocol used to generate BCOP data differed considerably from the other studies (e.g., a spectrophotometer was used to measure opacity instead of an opacitometer; solids were applied neat instead of as a 20% solution or suspension).

2.3 EU Ocular Hazard Classification System

The eight studies (Gautheron et al. [1994]; Balls et al. [1995]; Swanson et al. [1995]; Casterton et al. [1996]; Gettings et al. [1996]; Southee [1998]; Swanson and Harbell [2000], and Bailey et al. [2004]) contained BCOP test data on 203 substances, 158 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification as defined by the EU classification system (EU [2001])[5] (see **Appendix III-A**). Based on results from *in vivo*

[4] For the purpose of this accuracy analysis, *in vivo* rabbit study results were used to identify EPA Category I irritants (i.e., severe irritants); substances classified as EPA Category II, III, or IV were identified as nonsevere irritants.

[5] For the purpose of this accuracy analysis, *in vivo* rabbit study results were used to identify EU R41 irritants (i.e., severe irritants); substances classified as EU R36 or not classified were identified as nonsevere irritants.

Table III-3. Evaluation of the Performance of the BCOP Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EPA[1] Classification System, by Study and Overall

Data Source	Anal.[2]	N[3]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[4]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Gautheron et al. 1994 (new)[5]	IVIS	48/52	73[6]	35/48	71	5/7	73	30/41	31	5/16	94	30/32	27	11/41	29	2/7
Gautheron et al. 1994 (old)[5]	IVIS	12/52	75[6]	9/12	67	4/6	83	5/6	80	4/5	71	5/7	17	1/6	33	2/6
Balls et al. 1995 (new)[7]	IVIS	53/59	66[6]	35/53	72	13/18	63	22/35	50	13/26	82	22/27	37	13/35	28	5/18
Balls et al. 1995 (old)	IVIS	55/59	69[6]	38/55	75	15/20	66	23/35	56	15/27	82	23/28	34	12/35	25	5/20
Swanson et al. 1995 (new)	IVIS	8/20	88	7/8	100	6/6	50	1/2	86	6/7	100	1/1	50	1/2	0	0/6
Swanson et al. 1995 (old)	IVIS	9/20	89	8/9	100	6/6	67	2/3	86	6/7	100	2/2	33	1/3	0	0/6
Gettings et al. 1996 (new)	Perm	25/25	80	20/25	60	6/10	93	14/15	86	6/7	78	14/18	7	1/15	40	4/10
Gettings et al. 1996 (old)	Perm	25/25	80	20/25	60	6/10	93	14/15	86	6/7	78	14/18	7	1/15	40	4/10
Casterton et al. 1996 (new)	O/P	56/97	62	35/56	41	11/27	83	24/49	69	11/16	60	24/40	17	5/29	59	14/27
Casterton et al. 1996 (old)	O/P	55/97	64	35/55	42	11/26	83	24/49	69	11/16	62	24/39	17	5/29	58	15/26
Southee 1998 (new)	IVIS	14/16	64[6]	9/14	40	2/5	78	7/9	50	2/4	70	7/10	22	2/9	60	3/5
Southee 1998 (old)	IVIS	14/16	64[6]	9/14	50	3/6	75	6/8	60	3/5	67	6/9	25	2/8	50	3/6
Swanson & Harbell 2000 (new)[7]	IVIS	9/13	89	8/9	75	3/4	100	5/5	100	3/3	83	5/6	0	0/5	25	1/4
Swanson & Harbell 2000 (old)	IVIS	9/13	89	8/9	75	3/4	100	5/5	100	3/3	83	5/6	0	0/5	25	1/4

1997

Data Source	Anal.[2]	N[3]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[4]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Bailey et al. 2004 (new)	IVIS	13/16	92	12/13	0	0/1	100	12/12	-	0/0	92	12/13	0	0/12	100	1/1
Bailey et al. 2004 (old)	IVIS	16/16	94	15/16	67	2/3	100	13/13	100	2/2	93	13/14	0	0/13	33	1/3
Entire Data Set[8] (new)		143/203	79	113/143	75	30/40	81	83/103	60	30/50	89	83/93	19	20/103	25	10/40
Entire Data Set (old)		117/200	80	93/117	73	33/45	83	60/72	74	35/47	83	58/70	17	12/72	27	12/45

1998 [1]EPA = U.S. Environmental Protection Agency (EPA [1996]).
1999 [2]Anal. = Analytical method used to transform the sample data into BCOP classification. IVIS = *In Vitro* Irritancy Score developed by Gautheron et al.
2000 (1994). Perm = Permeability value only used to classify *in vitro* ocular irritancy in the BCOP assay; an OD_{490} value >0.600 was considered a severe irritant.
2001 O/P = irritation class based on the endpoint (opacity or permeability) with the highest score for its respective range (Casterton et al. 1996).
2002 [3]N = Number of substances included in this analysis/the total number of substances in the study.
2003 [4]Data used to calculate the percentage.
2004 [5]New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the previous analysis included in the draft BCOP BRD.
2005 [6]Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the multiple testing
2006 laboratories and tests (for substances tested multiple times in a laboratory).
2007 [7]The test substance ethanol was evaluated in two different *in vivo* studies (ECETOC [1998]; Swanson and Harbell [2000]), producing discordant results with
2008 respect to EPA classification; the analysis was performed using the Category I classification.
2009 [8]Data from Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000), and
2010 Bailey et al. (2004) were pooled together and an overall *in vitro* classification was assigned for each test substance based on the majority and/or most severe
2011 classification obtained across tests and testing laboratories. Data from Casterton et al. (1996) were not included in this analysis since the protocol used to
2012 generate BCOP data differed considerably from the other studies (e.g., a spectrophotometer was used to measure opacity instead of an opacitometer, and
2013 solids were applied neat instead of as a 20% solution or suspension).

rabbit eye experiments, 49 of the 158 substances were classified as severe irritants (i.e., Category I), the other 109 substances were classified as nonsevere irritants (either Category R36 or not classified). The 45 substances that could not be classified according to the EU classification system due to the lack of adequate animal data are so noted in **Appendix III-A**.

Based on the data provided in the eight studies, and when results across multiply tested substances were combined to generate a single consensus call per test substance, the BCOP test method has an accuracy of 68% to 92%, a sensitivity of 52% to 100%, a specificity of 64% to 100%, a false positive rate of 0% to 36%, and a false negative rate of 0% to 48% (**Table III-4**).

In terms of an overall accuracy analysis, combining the data from Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000), and Bailey et al. (2004), the BCOP test method has an accuracy of 80% (114/143), a sensitivity of 82% (33/40), a specificity of 79% (81/103), a false positive rate of 21% (22/103), and a false negative rate of 18% (7/40). The performance characteristics for the pooled studies are provided in **Table III-4**. Data from Casterton et al. (1996) were not included in the overall accuracy analysis since the protocol used to generate BCOP data differed considerably from the other studies (e.g., a spectrophotometer was used to measure opacity instead of an opacitometer; solids were applied neat instead of as a 20% solution or suspension).

2.4 Accuracy of the BCOP Test Method for the GHS Ocular Hazard Classification System, by Chemical Class and Property of Interest-Reanalysis

In order to further evaluate discordant responses of the BCOP test method relative to the *in vivo* hazard classification, several accuracy sub-analyses were performed. These included specific classes of chemicals with sufficiently robust numbers of substances ($n \geq 5$), as well as certain properties of interest considered relevant to ocular toxicity testing (e.g., pesticides, surfactants, pH, physical form). Because the international community will soon adopt the GHS classification system for hazard labeling (UN [2003]), and considering that there were only modest differences in overall BCOP test method accuracy among the three regulatory classification systems (i.e., EPA, EU, GHS), these sub-analyses were focused only on the GHS system.

As indicated in **Table III-5**, there were some notable trends in the performance of the BCOP test method among these subgroups of substances. The chemical class of substances that was most consistently overpredicted according the GHS classification system (i.e., were false positives[6]) by the BCOP test method is alcohols. Nine out the 19 overpredicted substances were alcohols. The remaining chemical classes represented among the overpredicted substances were carboxylic acids (3), ketones (3), heterocyclic compounds (2), esters (1), and hydrocarbons (1). Among the 35 substances labeled as surfactants only 5% (1/21) were overpredicted by the BCOP test method. The only overpredicted surfactant was a surfactant-containing formulation.

[6] False positive in this context refers to a substance that was classified as a severe ocular irritant by the BCOP test method, but as a nonsevere (mild or moderate) irritant or nonirritant based on *in vivo* data.

Table III-4. Evaluation of the Performance of the BCOP Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EU[1] Classification System, by Study and Overall

Data Source	Anal.[2]	N[3]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[4]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Gautheron et al. 1994 (new)[5,6]	IVIS	48/52	73[7]	35/48	71	5/7	73	30/41	31	5/16	94	30/32	27	11/41	29	2/7
Gautheron et al. 1994 (old)[5]	IVIS	51/52	75[7]	38/51	75	6/8	74	32/43	39	7/18	94	31/33	26	11/43	25	2/8
Balls et al. 1995 (new)	IVIS	50/59	68[7]	34/50	74	14/19	64	20/31	56	14/25	80	20/25	36	11/31	26	5/19
Balls et al. 1995 (old)	IVIS	59/59	71	42/59	76	16/21	68	26/38	55	16/29	83	25/30	34	13/38	24	5/21
Swanson et al. 1995 (new)	IVIS	9/20	89	8/9	100	6/6	67	2/3	86	6/7	100	2/2	33	1/3	0	0/6
Swanson et al. 1995 (old)	IVIS	9/20	78	7/9	100	5/5	50	2/4	71	5/7	100	2/2	50	2/4	0	0/5
Gettings et al. 1996 (new)	Perm	23/25	87	20/23	75	6/8	93	14/15	86	6/7	88	14/16	7	1/15	25	2/8
Gettings et al. 1996 (old)	Perm	25/25	80	20/25	67	4/6	84	16/19	57	4/7	89	16/18	16	3/19	33	2/6
Casterton et al. 1996 (new)	O/P	54/97	70	38/54	52	13/25	86	25/29	76	13/17	68	25/37	14	4/29	48	12/25
Casterton et al. 1996 (old)	O/P	60/97	73	44/60	54	13/24	86	31/36	72	13/18	74	31/42	14	5/36	46	11/24
Southee 1998 (new)	IVIS	14/16	79[7]	11/14	67	4/6	88	7/8	80	4/5	78	7/9	12	1/8	33	2/6
Southee 1998 (old)	IVIS	15/16	73[7]	11/15	60	3/5	80	8/10	60	3/5	80	8/10	20	2/10	40	2/5
Swanson & Harbell 2000 (new)	IVIS	9/13	78	7/9	100	1/1	75	6/8	33	1/3	100	6/6	25	2/8	0	0/1
Swanson & Harbell 2000 (old)	IVIS	9/13	78	7/9	100	1/1	75	6/8	33	1/3	100	6/6	25	2/8	0	0/1

2062

Data Source	Anal.[2]	N[3]	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
			%	No.[4]	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Bailey et al. 2004 (new)	IVIS	13/16	92	12/13	67	2/3	100	10/10	100	2/2	91	10/11	0	0/10	33	1/3
Bailey et al. 2004 (old)	IVIS	16/16	94	15/16	67	2/3	100	13/13	100	2/2	93	13/14	0	0/13	33	1/3
Entire Data Set (new)[8]		143/203	80	114/143	82	33/40	79	81/103	60	33/55	92	81/88	21	22/103	18	7/40
Entire Data Set (old)		157/200	77	121/157	78	31/40	77	90/117	55	33/60	91	88/97	23	27/117	23	9/40

2063 [1]EU = European Union (EU [2001]).
2064 [2]Anal. = Analytical method used to transform the sample data into BCOP classification. IVIS = *In Vitro* Irritancy Score developed by Gautheron et al.
2065 (1994). Perm = Permeability value only used to classify *in vitro* ocular irritancy in the BCOP assay; an OD_{490} value >0.600 was considered a severe irritant.
2066 O/P = irritation class based on the endpoint (opacity or permeability) with the highest score for its respective range (Casterton et al. [1996]).
2067 [3]N = Number of substances included in this analysis/the total number of substances in the study.
2068 [4]Data used to calculate percentage.
2069 [5]Accuracy analysis based on EEC (1984) classifications in Gautheron et al. (1994).
2070 [6]New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the previous analysis included in the Draft BCOP BRD.
2071 [7]Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the multiple testing
2072 laboratories and tests (for substances tested multiple times in a laboratory).
2073 [8]Data from Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000), and
2074 Bailey et al (2004) were pooled together and an overall *in vitro* classification was assigned for each test substance based on the majority and/or most severe
2075 classification obtained across tests and testing laboratories. Data from Casterton et al. (1996) were not included in this analysis since the protocol used to
2076 generate BCOP data differed considerably from the other studies (e.g., a spectrophotometer was used to measure opacity instead of an opacitometer, and
2077 solids were applied neat instead of as a 20% solution or suspension).

Table III-5. False Negative and False Positive Rates of the BCOP Test Method, by Chemical Class and Properties of Interest, for the GHS[1] Classification System

Category	N[2]	False Positive Rate[3]		False Negative Rate[4]	
		%	No.[5]	%	No.
Overall	147	20	21/104	16	7/43
Chemical Class[6]					
Alcohol	21	50	9/18	67	2/3
Amine/Amidine	8	0	0/4	0	0/4
Carboxylic acid	16	33	3/9	14	1/7
Ester	12	12	1/8	0	0/4
Ether/Polyether	6	0	0/5	0	0/1
Heterocycle	12	33	2/6	17	1/6
Hydrocarbon	11	9	1/11	-	0/0
Inorganic salt	5	0	0/3	0	0/2
Ketone	9	33	3/9	-	0/0
Onium compound	11	0	0/3	0	0/8
Properties of Interest					
Liquids	93	26	18/69	4	1/24
Solids	34	10	2/20	43	6/14
Pesticide	8	33	1/3	40	2/5
Surfactant – Total[7]	35	5	1/21	7	1/14
-nonionic	5	0	0/4	0	0/1
-anionic	3	0	0/2	100	1/1
-cationic	6	0	0/1	0	0/7
pH – Total[8]	24	-	-	21	5/24
- acidic (pH < 7.0)	11	-	-	18	2/11
- basic (pH > 7.0)	13	-	-	23	3/13
Category 1 Subgroup[9] - Total	38	-	-	18	7/38
- 4 (CO=4 at any time)	20	-	-	15	3/20
- 3 (severity/persistence)	1	-	-	0	0/1
- 2 (severity)	4	-	-	25	1/4
- 2-4 combined[10]	25	-	-	17	4/24
- 1 (persistence)	13	-	-	23	3/13

[1]GHS = Globally Harmonized System (UN [2003]).

[2]N = number of substances.

[3]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.

[4]False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.

[5]Data used to calculate the percentage.

[6]Chemical classes included in this table are represented by at least five substances tested in the BCOP test method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh) as defined in **Appendix B**.

[7]Combines single chemicals labeled as surfactants along with surfactant-containing formulations.

[8]Total number of GHS Category 1 substances for which pH information was obtained.

[9]NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1 substance. 1: based on lesions that are persistent; 2: based on lesions that are severe (not including Corneal Opacity [CO]=4); 3: based on lesions that are severe (not including CO=4) and persistent; 4: CO = 4 at any time.

[10]Subcategories 2 to 4 combined to allow for a direct comparison of GHS Category 1 substances classified *in vivo* based on some lesion severity component and those classified based on persistent lesions alone.

With regard to physical form of the substances overpredicted by the BCOP test method, 18 were liquids and two were solids. Considering the proportion of the total available database, liquids (93/127; 73%) appear more likely than solids (34/127; 27%) to be overpredicted by the BCOP test method

Although there were a relatively small number (4) of substances (i.e., were false negatives[7]), alcohols (2) were most often underpredicted by the BCOP test method according to the GHS classification system (see **Appendix III-A**). The other chemical classes represented were carboxylic acids (1) and heterocyclic compounds (1). As can be seen in **Table III-5**, the 35 substances labeled as surfactants were rarely underpredicted by the BCOP test method (7% [1/14] false negative rate). The only underpredicted surfactant was an anionic form. With regard to physical form of the substances underpredicted by the BCOP test method, 6 were solids and one was a liquid. Despite the proportion of the total available database, solids (34/127; 27%) appear more likely than liquids (93/127; 73%) to be underpredicted by the BCOP test method. There was no definitive difference among the underpredicted substances for which pH information was available, as two were acidic (pH < 7.0) and three were basic (pH > 7.0), and considering the comparable proportion of acidic and basic underpredicted substances (2/11; 18% vs. 3/13; 23%). Finally, the 38 underpredicted substances were more likely to be substances classified *in vivo* based on persistent lesions, rather than on severe lesions, as evidenced by an analysis of NICEATM-defined GHS Category 1 sub-groupings (**Table III-5**).

2.5 Accuracy of the BCOP Test Method for Identifying Ocular Corrosives and Severe Irritants – Summary of Reanalysis

As detailed in **Section III-1.0**, additional or new data relevant to the BCOP test method were received after the Expert Panel meeting on January 11 and 12, 2005 that increased the size of the comparative BCOP:*in vivo* rabbit eye test database from 120 to 147 substances for the GHS classification system (UN [2003]), 117-143 for the EPA classification system (EPA [1996]). Conversely, the size of the comparative BCOP:*in vivo* rabbit eye test database was decreased from 157 to 143 substances for the EU classification system (EU [2001]). As can be seen in **Tables III-2** through **III-4**, the overall accuracy stayed the same (draft BCOP BRD: 77-80%, depending on the classification system used; reanalysis: 80% for all classification systems). The false positive rate was reduced from 23% (draft BCOP BRD analysis) to 21% (reanalysis) for the EU classification system, but was increased from 17-19% (draft BRD BCOP analysis) to 19-20% (reanalysis) for the EPA and GHS classification systems, respectively, while the false negative rate was reduced for all three classification systems (from 23-27% [draft BCOP BRD analysis] to 16-25% [reanalysis]).

Similar to the original analysis, the revised analysis indicated that alcohols are often overpredicted (50% [9/18] false positive rate) in the BCOP test method. Carboxylic acids (3/9) and heterocyclic compounds (2/6) had a false negative rate of 33%.

[7] False negative in this context refers to a substance that was classified as a nonsevere (mild or moderate) irritant or nonirritant by the BCOP test method, but as a severe irritant based on *in vivo* data.

As noted in **Section III-2.4**, 18 of the 20 overpredicted substances were liquids while two were solids. Considering the proportion of the total available database, liquids (93) appear more likely than solids (34) to be overpredicted by the BCOP test method. In comparison to the original analysis, the overprediction of solid substances was reduced (from 44% [4/9] to 10% [2/20] false positive rate), while the false positive rate for liquids was increased from 21% (14/66) to 26% (18/69).

With regard to physical form of the substances underpredicted by the BCOP test method, six were solids and one was a liquid. Given the proportion of the total available database, solids (34/127; 27%) appear more likely than liquids (93/127; 73%) to be underpredicted by the BCOP test method. In comparison to the original analysis, the underprediction of solid substances was increased (from 31% [4/13] to 43% [6/14] false negative rate), while the false negative rate for liquids was reduced in the revised analysis from 18% (5/28) to 4% (1/24).

Using the expanded database, an analysis was conducted of the ability of the BCOP test method to identify ocular corrosives and severe irritants, depending on the nature of the *in vivo* ocular lesions (i.e., severity and/or persistence) responsible for classification of a substance as an ocular corrosive/severe irritant. As indicated in **Table III-5**, the 38 underpredicted substances were more likely to be substances classified *in vivo* based on persistent lesions (false negative rate = 23% [3/13]), rather than on severe lesions (false negative rate = 17% [4/24]).

A new analysis not included originally was an evaluation of accuracy related to acidic or basic pH. Among the five underpredicted substances for which pH information was available, 2 were acidic (pH < 7.0) and three were basic (pH > 7.0). Basic substances (13) occupy a comparable proportion of the total database to acidic substances (11), and therefore these differences do not appear to be significant. However, it is noted that pH information was obtained for only 28 of the 43 total Category 1 substances.

Table III-6 provides a summary of the revised analysis of the overall performance of the BCOP test method defined by the GHS classification system (UN [2003]). As noted from this analysis, the false positive substances were scattered among the three "nonsevere irritant" classifications (i.e., GHS Category 2A, 2B, or nonirritant). This includes nine (9/75) nonirritating substances that were classified as severe irritants by the BCOP. However, the mild irritants (Category 2B; n = 1/7) were less likely to be overpredicted as severe irritants/ocular corrosives than the moderate irritants (Category 2A, n = 11/22). The small number of false negative substances (7) was most often confined to those classified, based on BCOP test results, as moderate irritants (n=5) although two false negative substances were classified as mild irritants.

In the reanalysis, compared to the overall false positive rate for the BCOP test method (20%; 21/104) the false positive rate for alcohols is 50% (9/18). However, the revised analysis indicates that the false positive rate for ketones is smaller than originally determined (False positive rate: draft BCOP BRD analysis: 2/3, 67%; reanalysis: 3/9, 33%; Solid false positive). Likewise, the false positive rate for solids is smaller than

Table III-6. Overall Performance of the BCOP Test Method in the Predicting the Irritancy of a Substance as Defined by the GHS[1] Classification System

		***In Vitro* Classification (BCOP)**			
		Severe	**Moderate[2]**	**Mild**	***Total***
***In Vivo* Classification (GHS)**	**Category 1**	36	5	2	**43**
	Category 2A	11	7	4	**22**
	Category 2B	1	4	2	**7**
	Nonirritant	9	22	44	**75**
	Total	**57**	**38**	**52**	**147[3]**

[1]GHS = Globally Harmonized System (UN [2003]).
[2]*In vitro* classification of moderate also includes those substances classified as "nonsevere" in some BCOP studies.
[3]Thirty substances included in **Appendix III-A** had insufficient data with which to assign a precise GHS classification and therefore were not included in this table.

previously calculated (False positive rate: draft BCOP BRD analysis: 4/9, 44%; reanalysis: 2/20, 10%). Furthermore, the reanalysis indicated that the false negative rate of liquids was smaller than previously determined (draft BCOP BRD analysis: 5/28, 18%; reanalysis: 1/24, 4%). Based on the reanalysis, the false positive and false negative rates for identification of ocular corrosives/severe irritants among surfactants and surfactant containing formulations were 5% (1/21) and 7% (1/14), respectively.

3.0 RELIABILITY OF THE BCOP TEST METHOD - REANALYSIS

An assessment of test method reliability (intralaboratory repeatability and intra- and inter-laboratory reproducibility) is an essential element of any evaluation of the performance of an alternative test method (ICCVAM [2003]). Repeatability refers to the closeness of agreement between test results obtained within a single laboratory when the procedure is performed on the same substance under identical conditions within a given time period (ICCVAM [1997, 2003]). Intralaboratory reproducibility refers to the determination of the extent to which qualified personnel within the same laboratory can replicate results using a specific test protocol at different times. Interlaboratory reproducibility refers to the determination of the extent to which different laboratories can replicate results using the same protocol and test chemicals, and indicates the extent to which a test method can be transferred successfully among laboratories. A reliability assessment includes reviewing the rationale for selecting the substances used to evaluate test method reliability, a discussion of the extent to which the substances tested represent the range of possible test outcomes and the properties of the various substances for which the test method is proposed for use, and a quantitative and/or qualitative analysis of repeatability and intra- and inter-laboratory reproducibility. In addition, measures of central tendency and variation are summarized for historical control data (negative, vehicle, positive), where applicable.

As noted in the draft BCOP BRD, quantitative BCOP test method data were available for replicate corneas within individual experiments or for replicate experiments within an individual laboratory for three studies (Gettings et al. [1996]; Southee [1998]; data submission from Dr. Joseph Sina). Therefore, an evaluation of the repeatability and intralaboratory reproducibility of the BCOP test method could be conducted. Additionally, comparable BCOP data were available for multiple laboratories within each of three comparative validation studies (Gautheron et al. [1994]; Balls et al. [1995]; Southee [1998]), which allowed for an evaluation of the interlaboratory reproducibility of the BCOP test method.

3.1 Substances Used to Re-evaluate the Reliability of the BCOP Test Method

Intralaboratory reliability analyses were conducted on the data from Gettings et al. (1996), Southee (1998) and a data submission from Dr. Joseph Sina. For the Gettings et al. (1996) study, mean permeability data from three different experiments on the 25 surfactant-based formulations evaluated the CTFA Phase III study were obtained, as well as the mean permeability value for the three experiments, the standard deviation and the corresponding %CV values. Dr. Joseph Sina submitted a study of 43 substances, which included detailed BCOP data for replicate corneas. In the Southee (1998) study, 16 substances were evaluated in three laboratories multiple times (2 to 5 experiments) for a total of 122 tests.

Interlaboratory reliability analyses were conducted on the data from Gautheron et al. (1994), Balls et al. (1995), and Southee (1998). Gautheron et al. (1994) included 52 substances, including 22 liquids, 22 solids, and eight surfactants (both solids and liquids). Balls et al. (1995) included 60 substances (i.e., there were 52 different substances with four substances tested at two different concentrations and two substances tested at three concentrations, for a total of 60 possible ocular irritation outcomes). One substance (thiourea) was tested *in vitro* in the BCOP test method but, due to its excessive toxicity *in vivo*, was excluded from the comparison of *in vitro* and *in vivo* test results. As noted above, the Southee (1998) study included 16 substances evaluated in three laboratories multiple times.

3.2 Reanalysis of BCOP Test Method Intralaboratory Repeatability

Generally, analyses of intralaboratory repeatability have included approaches such as:

- a CV analysis, which is a statistical measure of the deviation of a variable from its mean (e.g., Holzhütter et al. [1996])
- ANOVA methods (e.g., Holzhütter et al. [1996]; ASTM [1999])

Two studies discussed in **Section 2.0** included intralaboratory repeatability data. For the Southee (1998) study, quantitative BCOP test method data were available for replicate corneas within individual experiments repeated two to five times for each test substance in three different laboratories. CV analyses were performed on within-experiment BCOP data, using the *In Vitro* Irritancy Score obtained for each test substance within each of the three testing laboratories. In addition, Dr. Joseph Sina submitted a study of 43 substances, which included detailed BCOP data for replicate corneas. A CV analysis was conducted

on the subset of substances provided by Dr. Sina that were tested using an incubation temperature of 32˚C, the recommended temperature for incubations in the proposed standardized protocol described in Appendix A of the draft BCOP BRD; substances incubated at room temperature were not included in this analysis. The updated information received subsequent to the release of the draft BCOP BRD did not affect these analyses and therefore these are not discussed again here (see the draft BCOP BRD, published November 1, 2004).

3.3 Reanalysis of BCOP Test Method Intralaboratory Reproducibility

Generally, analyses of intralaboratory reproducibility have included approaches such as:

- a CV analysis, which is a statistical measure of the deviation of a variable from its mean (e.g., Holzhütter et al. [1996])
- ANOVA methods (e.g., Holzhütter et al. [1996]; ASTM [1999])

Two of the studies discussed in **Section 2.0** included intralaboratory reproducibility data (Gettings et al. [1996]; Southee [1998]). For the Southee (1998) study, quantitative BCOP test method data were available for replicate corneas within individual experiments repeated two to five times for each test substance in three different laboratories. CV analyses were performed on between-experiment BCOP data, using the *In Vitro* Irritancy Score obtained for each test substance within each of the three testing laboratories. For the Gettings et al. (1996) study, Dr. John Harbell provided the mean permeability data obtained from three different experiments on the 25 surfactant-based formulations evaluated the Cosmetic, Toiletry, and Fragrance Association (CTFA) Phase III study, as well as the mean permeability value for the three experiments, the standard deviation and the corresponding %CV values. The updated information received subsequent to the release of the draft BCOP BRD did not affect these analyses and therefore these are not discussed again here (see the draft BCOP BRD, November 1, 2004).

3.4 Reanalysis of BCOP Test Method Interlaboratory Reproducibility

Generally, analyses of interlaboratory variability have included approaches such as:

- the extent of concordance among laboratories in assigning the same regulatory classification for a particular substance (e.g., Holzhütter et al. [1996])
- bivariant scatter diagrams/correlation analyses for pairs of laboratories to assess the extent possibility of divergence (e.g., Holzhütter et al. [1996])
- a CV analysis (e.g., Holzhütter et al. [1996])
- ANOVA methods (e.g., Holzhütter et al. [1996]; ASTM [1999])

Several of the studies discussed in **Section 2.0** included interlaboratory data for at least a subset of the substances evaluated. The ability of the BCOP test method to reproducibly identify ocular corrosives/severe irritants versus nonsevere irritants/nonirritants was evaluated using two approaches. While the draft BRD contained the same analysis as detailed below, new information regarding *in vivo* classification of substances according to the three regulatory classification schemes was provided, which resulted in changes to

the classification of some substances. Therefore, a revised analysis was conducted to reflect the updated classifications. However, while the tables include the comparative results from the original and revised analyses, the results discussed in the text pertain to the revised analysis only.

In the first approach, a qualitative assessment of reproducibility was conducted. In this evaluation, the individual laboratory *in vitro* ocular irritation classification for each substance was used to evaluate the extent of agreement among the participating laboratories in their ability to identify ocular corrosives/severe irritants versus nonsevere irritants/nonirritants. The reliability of BCOP was assessed separately for each study (i.e., publication). Substances classified, based on BCOP data, as corrosive/severe irritants or nonsevere irritants/nonirritants were further classified by their *in vivo* rabbit eye test results, as determined within the GHS, EPA, and EU classification schemes. Because the focus of this reliability assessment is on the interlaboratory reproducibility of BCOP in identifying corrosives/severe irritants versus nonsevere irritants/nonirritants, considerable variability could exist among laboratories in their classification of substances as nonsevere irritants or nonirritants (e.g., three laboratories could classify a chemical as a nonirritant and one laboratory could classify the same chemical as an moderate irritant; for this analysis this would be considered 100% agreement between laboratories) that would not be apparent from this analysis.

3.4.1 Qualitative Reanalysis of Interlaboratory Reproducibility

3.4.1.1 *GHS Ocular Hazard Classification System*

Reliability analyses for the BCOP test method were evaluated for the following three studies: Balls et al. (1995), Gautheron et al. (1994), and Southee (1998). The agreement of classification calls among participating laboratories and the relationship to the *in vivo* classification (UN [2003]) for the substances tested in each validation in each study is provided in **Table III-7**.

For the study by Balls et al. (1995), the five participating laboratories were in 100% agreement in regard to the ocular irritancy classification for 41 (68%) of the 60 substances tested. The extent of agreement between testing laboratories was the same for substances identified from *in vivo* rabbit eye data as corrosives/severe irritants or as nonsevere irritants/nonirritants (76% of the accurately identified severe and nonsevere substances were shown to have 100% classification agreement among testing laboratories). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives (i.e., positive *in vitro* but negative *in vivo*). For instance, 63% of the false positives exhibited less than 100% agreement in the irritancy classifications among laboratories.

For the study by Gautheron et al. (1994), there was 100% agreement in regard to the ocular irritancy classification for 35 (67%) of the 52 substances, which were tested in either 11 or 12 laboratories. Discordance in the classification results was present for

Table III-7. Evaluation of the Reliability of the BCOP Test Method in Predicting Ocular Corrosives and Severe Irritants as Defined by the GHS[1] Classification System, by Study

Report	Classification (*In Vivo/In Vitro*)[2]	No. of Testing Labs	N[3]	Substances with 100% Agreement among Labs[4]	Substances with 91-92% Agreement among Labs	Substances with 82-83% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 73% Agreement among Labs	Substances with 64-67% Agreement among Labs	Substances with 58-60% Agreement among Labs	Substances with ≤ 55% Agreement among Labs
Balls et al. (1995)	+/+ (new)[5]	5	17	13 (76%)			3 (18%)			1 (6%)	
	+/+ (old)[5]	5	17	14 (82%			2 (12%)			1 (6%)	
	+/- (new)	5	5	3 (60%)			1 (20%)			1 (20%)	
	+/- (old)	5	5	3 (60%)			1 (20%)			1 (20%)	
	-/+ (new)	5	11	4 (36%)			4 (36%)			3 (27%)	
	-/+ (old)	5	12	4 (33%)			5 (42%)			3 (25%)	
	-/- (new)	5	21	16 (76%)			2 (10%)			3 (14%)	
	-/- (old)	5	23	17 (74%)			2 (9%)			4 (17%)	
	?/- (new)	5	4	3 (75%)						1 (25%)	
	?/- (old)	5	2	2 (100%)						0 (0%)	
	?/+ (new)	5	2	2 (100%)							
	?/+ (old)	5	1	1 (100%)							
	Total (new)		60	41 (68%)			10 (17%)			9 (15%)	
	Total (old)		60	41 (68%)			10 (17%)			9 (15%)	
Gautheron et al. (1994)	+/+ (new)	11	5	3 (60%)		1 (10%)					1 (10%)
		12	1	1(100%)							
	+/+ (old)	11	4	2 (50%)		1 (25%)					1 (25%)
		12	1	1 (100%)							
	+/- (new)	11	1			1(100%)					
		12	1	1(100%)							
	+/- (old)	11	1			1 (100%)					
		12	1	1 (100%)							
	-/+ (new)	11	4	2 (50%)		1 (25%)		1 (25%)			
		12	5	2 (40%)	1 (20%)						2 (40%)
	-/+ (old)	11	1						1 (100%)		
	-/- (new)	11	2			1 (50%)			1 (50%)		
		12	28	23 (81%)	1 (4%)	3 (11%)			1 (4%)		
	-/- (old)	11	4	3 (75%)			1 (25%)				
		12	1			1 (100%)					
	?/- (new)	11	1					1 (100%)			
		12	1	1 (100%)							
	?/- (old)	11	11	8 (73%)		2 (18%)		1 (9%)			

Report	Classification (*In Vivo/In Vitro*)[2]	No. of Testing Labs	N[3]	Substances with 100% Agreement among Labs[4]	Substances with 91-92% Agreement among Labs	Substances with 82-83% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 73% Agreement among Labs	Substances with 64-67% Agreement among Labs	Substances with 58-60% Agreement among Labs	Substances with ≤ 55% Agreement among Labs
		12	16	15 (94%)	1 (6%)						
	?/+ (new)	11	3	1 (33%)	1 (33%)				1 (33%)		
	?/+ (old)	11	7	4 (57%)	1 (14%)	1 (14%)		1 (14%)			
		12	4	2 (50%)	1 (25%)					1 (25%)	
	Total (new)		52	34 (65%)	3 (6%)	7 (13%)		2 (4%)	3 (6%)		3 (6%)
	Total (old)		51	36 (71%)	3 (6%)	6 (12%)	1 (2%)	2 (4%)	1 (2%)	1 (2%)	1 (2%)
Southee (1998)	+/+ (new)	3	4	4 (100%)							
	+/+ (old)	3	3	3 (100%)							
	+/- (new)	3	3	3 (100%)							
	+/- (old)	3	3	3 (100%)							
	-/+ (new)	3	1	1 (100%)							
	-/+ (old)	3	2	2 (100%)							
	-/- (new)	3	7	6 (86%)					1 (14%)		
	-/- (old)	3	6	5 (83%)					1 (17%)		
	?/- (new)	3	1	1 (100%)							
	?/- (old)	3	2	2 (100%							
	?/+ (new)	-	0								
	?/+ (old)	-	0								
	Total (new)		16	15 (94%)					1 (6%)		
	Total (old)		16	15 (94%)					1 (6%)		

[1]GHS = Globally Harmonized System (UN [2003]).

[2]A "+" indicates that the substance was assigned an overall classification of corrosive or a severe irritant (Category 1); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category 2A, 2B) or nonirritant; a "?" indicates that, due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects; insufficient dose volume), a GHS classification could not be made. See **Section 2.0** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.

[3]N = number of substances.

[4]Number in parentheses indicates percentage of tested chemicals.

[5]New = accuracy statistics based on revised analysis; Old = accuracy statistics based on the previous analysis included in the draft BCOP BRD.

substances that were correctly identified as corrosives/severe irritants and as nonsevere irritants/nonirritants. For the study by Southee (1998), there was 100% agreement in regard to the ocular irritancy classification for 15 (94%) of the 16 substances. Discordance in the classification results was present for only one substance that was correctly identified as a nonsevere irritant/nonirritant.

3.4.1.2 *EPA Ocular Hazard Classification System*

Reliability analyses for the BCOP test method were evaluated for the following three studies: Balls et al. (1995), Gautheron et al. (1994), and Southee (1998). The agreement of classification calls among participating laboratories and its relationship to the *in vivo* classification (EPA [1996]) for the substances tested in each validation in each study is provided in **Table III-8**.

The participating laboratories of Balls et al. (1995) were in 100% agreement in regard to the ocular irritancy classification for 40 (67%) of the 60 substances tested. The agreement among laboratories was greatest for accurately identified corrosives/severe irritants when compared to any other combination of *in vivo* and *in vitro* results (77% of the accurately identified corrosives/severe irritants exhibited 100% classification agreement among laboratories). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives. For instance, 61% of the false positives exhibited less than 100% agreement among laboratories in the irritancy classifications.

The participating laboratories of Gautheron et al. (1994) were in 100% agreement in regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) for 36 (71%) of the 51 tested substances. Discordant results were observed for substances that were correctly identified as corrosive/severe irritant or nonsevere/irritant/nonirritant, as well as for false negatives and false positives.

For the report by Southee (1998), there was 100% agreement in regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) for 15 (94%) of the 16 substances. Discordance in the classification results was present for only one substance that was correctly identified as a nonsevere irritant/nonirritant.

3.4.1.3 *EU Ocular Hazard Classification System*

Reliability analyses for the BCOP test method were evaluated for the following three studies: Balls et al. (1995), Gautheron et al. (1994), and Southee (1998). The agreement of classification calls among participating laboratories and its relationship to the *in vivo* classification (EU [2001]) for the substances tested in each validation in each study is provided in **Table III-9**.

The participating laboratories were in 100% agreement in regard to the ocular irritancy classification for 40 (67%) of the 60 substances tested by Balls et al. (1995). The extent of agreement among laboratories was greatest for accurately identified corrosives/severe irritants when compared to any other combination of *in vivo* and *in vitro* results (86% of the accurately identified corrosives/severe irritants exhibited 100% classification agreement

Table III-8. Evaluation of the Reliability of the BCOP Test Method In Predicting Ocular Corrosives and Severe Irritants as Defined by the EPA[1] Classification System, by Study

Report	Classification (*In Vivo/In Vitro*)[2]	No. of Testing Labs	N[3]	Substances with 100% Agreement among Labs[4]	Substances with 91-92% Agreement among Labs	Substances with 82-83% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 73% Agreement among Labs	Substances with 64-67% Agreement among Labs	Substances with 58-60% Agreement among Labs	Substances with ≤ 55% Agreement among Labs
Balls et al. (1995)	+/+ (new)[5]	5	13	10 (77%)			2 (15%)			1 (8%)	
	+/+ (old)[5]	5	17	14 (82%)			2 (12%)			1 (6%)	
	+/- (new)	5	5	3 (60%)			1 (20%)			1 (20%)	
	+/- (old)	5	5	3 (60%)			1 (20%)			1 (20%)	
	-/+ (new)	5	13	5 (38%)			5 (38%)			3 (23%)	
	-/+ (old)	5	12	4 (33%)			5 (42%)			3 (25%)	
	-/- (new)	5	22	15 (68%)			4 (33%)			3 (25%)	
	-/- (old)	5	23	17 (74%)			2 (9%)			4 (17%)	
	?/- (new)	5	3	3 (100%)							
	?/- (old)	5	2	2 (100%)							
	?/+ (new)	5	4	4 (100%)							
	?/+ (old)	5	1	1 (100%)							
	Total (new)		60	40 (67%)			12 (20%)			8 (13%)	
	Total (old)		60	41 (68%)			10 (17%)			9 (15%)	
Gautheron et al. (1994)	+/+ (new)	11	4	2 (50%)		1 (25%)					1 (25%)
		12	1	1 (100%)							
	+/+ (old)	11	2			1 (33%)					1 (33%)
		12	1	1 (100%)							
	+/- (new)	11	1			1 (100%)					
		12	1	1 (100%)							
	+/- (old)	11	1			1 (100%)					
		12	1	1 (100%)							
	-/+ (new)	11	5	3 (60%)				1 (20%)	1 (20%		
		12	5	2 (40%)	1 (20%)					1 (20%)	1 (20%)
	-/+ (old)	11	1						1 (100%)		
	-/- (new)	11	11	8 (73%)		2 (18%)			1 (9%)		
		12	19	17 (90%)	1 (5%)	1 (5%)					
	-/- (old)	11	4	3 (75%)			1 (25%)				
		12	1			1 (100%)					
	?/- (new)	11	1					1 (100%)			
		12	1	1 (100%)							
	?/- (old)	11	11	8 (73%)		2 (18%)		1 (9%)			
		12	16	15 (94%)	1 (6%)						

Report	Classification (*In Vivo/In Vitro*)[2]	No. of Testing Labs	N[3]	Substances with 100% Agreement among Labs[4]	Substances with 91-92% Agreement among Labs	Substances with 82-83% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 73% Agreement among Labs	Substances with 64-67% Agreement among Labs	Substances with 58-60% Agreement among Labs	Substances with ≤ 55% Agreement among Labs
	?/+ (new)	11	2	1 (50%)	1 (50%)						
	?/+ (old)	11	9	6 (57%)	1 (14%)	1 (14%)		1 (14%)			
		12	4	2 (50%)	1 (25%)					1 (25%)	
	Total (new)		51	36 (71%)	3 (6%)	5 (10%)		2 (4%)	2 (4%)	1 (2%)	2 (4%)
	Total (old)		51	36 (71%)	3 (6%)	6 (12%)	1 (2%)	2 (4%)	1 (2%)	1 (2%)	1 (2%)
Southee (1998)	+/+ (new)	3	2	2 (100%)							
	+/+ (old)	3	3	3 (100%)							
	+/- (new)	3	3	3 (100%)							
	+/- (old)	3	3	3 (100%)							
	-/+ (new)	3	2	2 (100%)							
	-/+ (old)	3	2	2 (100%)							
	-/- (new)	3	7	6 (86%)					1 (14%)		
	-/- (old)	3	6	5 (83%)					1 (17%)		
	?/- (new)	3	1	1 (100%)							
	?/- (old)	3	2	2 (100%)							
	?/+ (new)	3	1	1 (100%)							
	?/+ (old)	-	0								
	Total (new)		16	15 (94%)					1 (6%)		
	Total (old)		16	15 (94%)					1 (6%)		

[1]EPA = U.S. Environmental Protection Agency (EPA [1996]).

[2]A "+" indicates that the substance was assigned an overall classification of corrosive or a severe irritant (Category I); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category II, III) or nonirritant (category IV); a "?" indicates that, due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects; insufficient dose volume), an EPA classification could not be made. See **Section 2.0** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.

[3]N = number of substances.

[4]Number in parentheses indicates percentage of tested chemicals.

[5]New = accuracy statistics based on revised analysis; Old = accuracy statistics based on the previous analysis included in the draft BCOP BRD.

among laboratories). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives, false negatives, and those substances accurately classified as nonsevere irritants/nonirritants. For instance, 63% of the false positives exhibited less than 100% agreement among laboratories in irritancy classifications.

The participating laboratories in Gautheron et al. (1994) were in 100% agreement in regard to the ocular irritancy classification for 36 (69%) of the 52 tested substances. Discordant results were observed for substances that were correctly identified as corrosive/severe irritant or nonsevere/irritant/nonirritant, as well as for false negatives and false positives.

For the report by Southee (1998), there was 100% agreement in regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) for 15 (94%) of the 16 substances.

3.4.2 Quantitative Reanalysis of Interlaboratory Reproducibility

As detailed in the draft BCOP BRD, to provide a quantitative assessment of interlaboratory variability, individual laboratory BCOP test results were used to calculate a mean and CV for the *In Vitro* Irritancy Score for each substance tested in Gautheron et al. (1994), Balls et al. (1995) and Southee (1998). Although a wide range of CV values were noted, mean and median CV values for the Balls et al. (1995) and the Southee (1998) study were less than 35%. These values were higher for the Gautheron et al. (1994) study (168% and 47%, respectively), although lower values were noted for substances predicted as severe irritants/corrosives in the BCOP test method (36% and 17% for mean and median CV values). The additional information received subsequent to the release of the draft BCOP BRD (November 1, 2004) did not affect these analyses, and therefore a reanalysis was not necessary.

3.4.3 Additional Reanalyses of Interlaboratory Reproducibility

As described in the draft BCOP BRD, Gautheron et al. (1994) found that 82.7% of the substances tested were classified the same by all laboratories when using a three-category system (i.e., mild irritant (BCOP score [0-25], moderate irritant [25.1-55] and severe irritant [≥55.1]). Also described in the draft BCOP BRD is the analysis of Balls et al. (1995), in which the interlaboratory correlation of BCOP results (permeability value, opacity value, and *In Vitro* Irritancy Score) generated from the five laboratories that participated this study was determined. This analysis yielded a wide range of correlation coefficients for the subsets of test substances.

The additional information received subsequent to the release of the draft BCOP BRD (November 1, 2004) did not affect these analyses, and therefore a reanalysis was not necessary.

2469 **Table III-9. Evaluation of the Reliability of the BCOP Test Method In Predicting Ocular Corrosives and Severe Irritants as**
2470 **Defined by the EU[1] Classification System, by Study**
2471

Report	Classification (*In Vivo/In Vitro*)[2]	No. of Testing Labs	N[3]	Substances with 100% Agreement among Labs[4]	Substances with 91-92% Agreement among Labs	Substances with 82-83% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 73% Agreement among Labs	Substances with 64-67% Agreement among Labs	Substances with 58-60% Agreement among Labs	Substances with ≤ 55% Agreement among Labs
Balls et al. (1995)	+/+ (new)[5]	5	14	12 (86%)			2 (14%)				
	+/+ (old)[5]	5	16	14 (88%)			2 (12%)				
	+/- (new)	5	5	2 (40%)			1 (20%)			2 (40%)	
	+/- (old)	5	5	3 (60%)			1 (20%)			1 (20%)	
	-/+ (new)	5	11	4 (36%)			4 (36%)			3 (27%)	
	-/+ (old)	5	13	4 (31%)			5 (38%)			4 (31%)	
	-/- (new)	5	20	15 (75%)			2 (10%)			3 (15%)	
	-/- (old)	5	25	19 (76%)			2 (8%)			4 (16%)	
	?/- (new)	5	5	5 (100%)							
	?/- (old)	-	0								
	?/+ (new)	5	5	3 (60%)			1 (20%)			1 (20%)	
	?/+ (old)	5	1	1 (100%)							
	Total (new)		60	40 (67%)			10 (17%)			9 (15%)	
	Total (old)		60	41 (68%)			10 (17%)			9 (15%)	
Gautheron et al. (1994)	+/+ (new)	11	5	3 (60%)		1 (20%)					1 (20%)
		12	1	1 (100%)							
	+/+ (old)	11	5	3 (60%)		1 (20%)					1 (20%)
		12	1	1 (100%)							
	+/- (new)	11	1			1 (100%)					
		12	1	1 (100%)							
	+/- (old)	11	1			1 (100%)					
		12	1	1 (100%)							
	-/+ (new)	11	5	2 (40%		1 (20%)		1 (20%)	1 (20%)		
		12	5	2 (40%)	1 (20%)					1 (20%)	1 (20%)
	-/+ (old)	11	4	1 (25%)	1 (25%)	1 (25%)		1 (25%)			
		12	5	2 (40%)	1 (20%)				1 (20%)	1 (20%)	
	-/- (new)	11	15	12 (80%)		2 (13%)			1 (7%)		
		12	15	13 (87%)	1 (7%)	1 (7%)					
	-/- (old)	11	15	11 (73%)		2 (13%)		1 (7%)	1 (7%)		
		12	17	15 (88%)	1 (6%)	1 (6%)					
	?/- (new)	11	1					1 (100%)			
		12	1	1 (100%)							
	?/- (old)	11	1	1 (100%)							

Report	Classification (*In Vivo/In Vitro*)[2]	No. of Testing Labs	N[3]	Substances with 100% Agreement among Labs[4]	Substances with 91-92% Agreement among Labs	Substances with 82-83% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 73% Agreement among Labs	Substances with 64-67% Agreement among Labs	Substances with 58-60% Agreement among Labs	Substances with ≤ 55% Agreement among Labs
	?/+ (new)	11	2	1 (50%)	1 (50%)						
	?/+ (old)	-	0								
	Total (new)		52	36 (69%)	3 (6%)	6 (12%)		2 (4%)	2 (4%)	1 (2%)	2 (4%)
	Total (old)		50	35 (70%)	3 (6%)	6 (12%)	1 (2%)	2 (4%)	2 (2%)	1 (2%)	1 (2%)
Southee (1998)	+/+ (new)	3	4	4 (100%)							
	+/+ (old)	3	3	3 (100%)							
	+/- (new)	3	2	2 (100%)							
	+/- (old)	3	2	2 (100%)							
	-/+ (new)	3	1	1 (100%)							
	-/+ (old)	3	2	2 (100%)							
	-/- (new)	3	7	6 (86%)					1 (14%)		
	-/- (old)	3	8	7 (88%)					1 (12%)		
	?/- (new)	3	2	2 (100%)							
	?/- (old)	3	1	1 (100%)							
	?/+ (new)	-	0								
	?/+ (old)	-	0								
	Total (new)		16	15 (94%)					1 (6%)		
	Total (old)		16	15 (94%)					1 (6%)		

[1]EU = European Union (EU [2001]).

[2]A "+" indicates that the substance was assigned an overall classification of corrosive or a severe irritant (Category I); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category II, III) or nonirritant (category IV); a "?" indicates that, due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects; insufficient dose volume), an EPA classification could not be made. See **Section 2.0** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.

[3]N indicates number of substances.

[4]Number in parentheses indicates percentage of tested chemicals.

[5]New = accuracy statistics based on revised analysis; Old = accuracy statistics based on the previous analysis included in the draft BCOP BRD.

3.5 BCOP Test Method Historical Positive and Negative Control Data - Reanalysis

An example of historical data for positive controls was provided by IIVS (current as of July 22, 2004), and is provided in the draft BCOP BRD.

3.6 Reliability of the BCOP Test Method for Identifying Ocular Corrosives and Severe Irritants – Summary of Reanalysis

As described in the draft BCOP BRD, a quantitative assessment of intralaboratory data (*In Vitro* Irritancy Scores) from two studies (Southee [1998]; Sina submission) was conducted for substances predicted as severe eye irritants. For the 16 substances evaluated in the Southee (1998) study, the median %CV for *In Vitro* Irritancy Scores for replicate corneas ranged from 11.8 to 14.2 for the three laboratories. For the 29 substances evaluated by Dr. Sina, the within experiment mean and median %CV values for *In Vitro* Irritancy Scores were 71 and 35, respectively. The dataset provided by Dr. Sina included 10 substances with low *In Vitro* Irritancy Scores, contributing to the increased variability of this dataset. However, the range of %CV values for the five substances predicted as severe irritants (*In Vitro* Scores >55.1) in this study is 1.1 to 13.

Also described in the draft BCOP BRD is a quantitative assessment of intralaboratory data (*In Vitro* Irritancy Scores) from two studies (Gettings et al. [1996]; Southee [1998]). For the Gettings et al. (1996) study, the between experiment (n=3) mean and median %CV values for permeability values were 33.4 and 29.0, respectively, for 25 surfactant-based personal care cleaning formulations. For the Southee (1998) study, the between experiment %CV values of *In Vitro* Irritancy Scores for the 16 substances tested two or more times in Laboratory 1, Laboratory 2, and Laboratory 3 was less than 35%. The mean %CV values for this study ranged from 12.6 to 14.8 for the three laboratories, while the median %CV values ranged from 6.7 to 12.4.

These analyses of intralaboratory reliability were not affected by the information received subsequent to the release of the draft BCOP BRD (November 1, 2004). However, the previous analysis also included an evaluation of interlaboratory reproducibility using both qualitative and quantitative approaches. While the quantitative analysis was unaffected by the new information that was received, the qualitative analysis (correct classification as an ocular corrosive/severe irritant or as a non-corrosive/nonsevere irritant) of the data provided for multiple laboratories in three studies (Gautheron et al. [1994]; Balls et al. [1995]; Southee [1998]) mandated that this analysis be repeated. The results for this analysis are presented in **Tables III-7** to **III-9**. The five participating laboratories for the Balls et al. (1995) study were in 100% agreement in regard to the ocular irritancy classification for 40 (67%) of the 60 substances tested *in vitro* in the study. In general, the extent of agreement between testing laboratories was greatest for substances identified from *in vivo* rabbit eye data as corrosives or severe irritants when compared to any other combination of *in vivo* and *in vitro* results (76% to 86%, depending on the classification system used, of the accurately identified severe substances were shown to have 100% classification agreement among testing laboratories). For the study by Gautheron et al. (1994), there was 100% agreement in regard to the ocular

irritancy classification for 35 to 36 (67% to 69%) of the 52 substances, which were tested in either 11 or 12 laboratories. Finally, for the study by Southee (1998), there was 100% agreement in regard to the ocular irritancy classification for 15 (94%) of the 16 substances.

4.0. REFERENCES

ASTM. 1999. Standard practice for conducting an interlaboratory study to determine the precision of a test methods. ASTM E691-92. In: Annual Book of ASTM Standards. Philadelphia, PA: American Society for Testing and Materials.

Bailey PT, Freeman JJ, Phillips RD, Merrill JC. 2004. Validation of the BCOP assay as a predictor of ocular irritation of various petrochemical products. Poster presentation at the Society of Toxicology 2004 meeting.

Balls M, Botham PA, Bruner LH, Spielmann H. 1995. The EC/HO international validation study on alternatives to the Draize eye irritation test. Toxicol In Vitro 9:871-929.

Casterton PL, Potts LF, Klein BD. 1996. A novel approach to assessing eye irritation potential using the bovine corneal opacity and permeability assay. J Toxicol - Cut and Ocular Toxicol 15:147-163.

EPA. 1996. Label Review Manual: 2nd Ed. EPA737-B-96-001. Washington, DC: U.S. Environmental Protection Agency.

EU. 2001. Commission Directive 2001/59/EC of 6 August 2001 adapting to technical progress for the 28th time Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labeling of dangerous substances. Official Journal of the European Communities L255:1-333.

Gautheron P, Giroux J, Cottin M, Audegond L, Morilla A, Mayordomo-Blanco L, Tortajada A, Haynes G, Vericat JA, Pirovano R, Tos EG, Hagemann C, Vanparys P, Deknudt G, Jacobs G, Prinsen M, Kalweit S, Spielmann H. 1994. Interlaboratory assessment of the bovine corneal opacity and permeability (BCOP) assay. Toxicol In Vitro 8:381-392.

Gettings S, Lordo R, Hintze K, Bagley D, Casterton P, Chudkowski M, Curren RD, Demetrulias JL, Dipasquale LC, Earl LK, Feder PI, Galli CL, Glaza SM, Gordon VC, Janus J, Kurtz PJ, Marenus KD, Moral J, Pape WJW, Renskers KJ, Rheins LA, Roddy MT, Rozen MG, Tedeschi JP, Zyracki J. 1996. The CTFA evaluation of alternatives program: An evaluation of in vitro alternatives to the Draize primary eye irritation test. (Phase III) surfactant-based formulations. Food Chem Toxic 34:79-117.

Holzhütter HG, Archer G, Dami N, Lovell DP, Saltelli A, Sjöström M. 1996. Recommendations of the application of biostatistical methods during the development and validation of alternative toxicological methods. ECVAM Biostatistics Task Force Report 1. ATLA 24:511-530.

ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods. NIH Publication No: 03-4508. Research Triangle Park, NC: National Toxicology Program. Available: http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm [accessed 12 July 2005].

ICCVAM. 1997. Validation and Regulatory Acceptance of Toxicological Methods: A Report of the *ad hoc* Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication No: 97-3981. Research Triangle Park, NC: National Toxicology Program. Available: http://iccvam.niehs.nih.gov/docs/guidelines/validate/pdf [accessed 12 July 2005].

MeSH Medical Subject Heading. Available: http://www.nlm.nih.gov/mesh [accessed 12 July 2005].

NICEATM. 2004. Draft Background Review Document. Current Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants: The Bovine Corneal Opacity and Permeability (BCOP) Test Method. Available: http://iccvam.niehs.nih.giv/methods/ocudocs/ocu_brd.htm [accessed 12 July 2005].

NIEHS. 2005. Second request for data on chemicals evaluated by *in vitro* or *in vivo* ocular irritancy test methods. Fed Reg 70:9661-9662. Available: http://iccvam.niehs.nih.gov/methods/eyeirrit.htm [accessed 12 July 2005].

NIEHS. 2004. Request for public comment on the nomination for ocular toxicity test methods and related activities and request for data on chemicals evaluated by *in vitro* or *in vivo* ocular irritancy test methods. Fed Reg 69:13859-13861. Available: http://iccvam.niehs.nih.gov/methods/eyeirrit.htm [accessed 12 July 2005].

Southee JA. 1998. Evaluation of the Prevalidation Process. Part 2, final report. Volume 2. The Bovine Corneal Opacity and Permeability (BCOP) Assay. European Community contract no. 11279-95-10F 1ED ISP GB.

Swanson JE, Harbell JW. 2000. Evaluating the eye irritancy potential of ethanolic test materials with the bovine corneal opacity and permeability assay. The Toxicologist 54:188-189.

Swanson JE, Lake LK, Donnelly TA, Harbell JW, Huggins J. 1995. Prediction of ocular irritancy of full-strength cleaners and strippers by tissue equivalent and bovine corneal assays. J Toxicol - Cut and Ocular Toxicol 14:179-195.

UN. 2003. Globally Harmonized System of Classification and Labelling of Chemicals (GHS). New York & Geneva: United Nations Publications. Available: http://www.unece.org/trans/danger/publi/ghs/ghs_rev00/00files_ehtml. [accessed 12 July 2005].

[This Page Intentionally Left Blank]